

PROCESS FOR PRODUCTION OF POLYPEPTIDES

5

Background of the Invention

Related Applications

10 This application is a non-provisional application filed under 37 CFR 1.53(b)(1), claiming priority under 35 USC 119(e) to provisional application number 60\274,384 filed March 9, 2001, the contents of which are incorporated herein by reference.

Field of the Invention:

15 This invention relates to improved vectors and methods for producing polypeptides using such vectors. In particular, this invention is related to improved expression of polypeptides from nucleic acids such as cloned genes and production of various polypeptides and proteins, including those of eukaryotic origin in prokaryotic hosts.

20 Description of Related Art

25 The level of production of a protein in a host cell is governed by three major factors: the number of copies of its gene within the cell, the efficiency with which those gene copies are transcribed, and the efficiency with which the resultant messenger RNA (mRNA) is translated. The quality of protein produced is similarly governed by various factors, including the anti-termination mechanism in the host cell.

30 Recombinant proteins produced in *E. coli* occasionally contain structural modifications that restrict their usefulness as therapeutic drugs or reagents for structure-function relationship studies. Such modifications include N- and C-terminal truncations, extensions, incomplete removal of N-terminal initiator methionine, misincorporation of lysine for arginine, and norleucine for methionine. For example, during the purification of recombinant murine interleukin-6 from *E. coli*, it was observed that 5-10% of the mIL-6 molecules contained a novel C-terminal modification (Tu et al., *J. Biol. Chem.*, 270: 9322-9326 (1995)).

35 This C-terminal "tag" is encoded by a small metabolically stable RNA of *E. coli* (10Sa RNA) (Chauhan and Apirion, *Mol. Microbiol.*, 3: 1481-1485 (1989)). 10Sa RNA, also known as transfer-messenger RNA, or tmRNA, contains a tRNA-like structure *in vivo* with the 5'- and 3'-end sequences and an internal reading frame encoding a "tag" peptide.

The primary cause of the production of truncated 10Sa-tagged proteins is the translation of mRNA truncated within the coding region (Keiler et al., Science, 271: 990-993 (1996)). Premature transcription termination and RNase cleavage appear to be the major factors capable of producing such truncated mRNA. The first of these factors, premature transcription termination, is potentially amenable to some type of transcription anti-termination. Several of these systems have been described, including  $\lambda$ N,  $\lambda$ Q, HK022, *rrn*, and *Psu* (Weisberg et al., J. Bacteriol., 181: 359-367 (1999)). Most of these systems are used to control gene expression temporally in phage development by transcribing through intergenic transcription terminators. The function of the *rrn* anti-termination is somewhat different, and it has been proposed to prevent *rho*-dependent transcription termination within the non-translated ribosomal RNA operons.

Despite the accumulation of considerable knowledge of these systems over many years, their only demonstrated usefulness in terms of recombinant technology has been in the control of gene expression by overriding intergenic transcriptional terminators (Mertens et al., Bio/Technol., 13: 175-179 (1995)). Other reports describe the failure of one of these systems (*rrn*) to alleviate problems within a translated coding sequence containing extensive secondary structure (Makrides, Microbiol. Rev., 60: 512-538 (1996)).

Several fusions of a protein with at least a portion of an anti-terminator protein have been disclosed, especially the N gene protein and most particularly the N-terminal fragment thereof (JP 9059299 published March 4, 1997; WO 89/03886 published May 5, 1989; WO 88/06628 published September 7, 1988; U.S. Pat. No. 5,834,184 issued November 10, 1998; EP 700,997 published March 13, 1996; U.S. Pat. No. 5,354,846 issued October 11, 1994; U.S. Pat. No. 5,618,715 issued April 8, 1997; U.S. Pat. No. 5,374,520 issued December 20, 1994; Zhukovskaya et al., Nucl. Acids. Res., 20: 6081-6090 (1992); Horiuchi et al., Biotechnol. Lett., 16: 113-118 (1994); Kamasawa et al., IFAC Symp. Ser., 10: 255-258 (1992); Kovgan et al., Vopr. Virusol., 31: 485-489 (1986)).

Several plasmids that contain an N utilization site for binding anti-terminator N protein produced by the host cell, such as *E. coli*, have been constructed (U.S. Pat. No. 5,256,546 issued October 26, 1993; EP 691,406 published January 10, 1996; U.S. Pat. No. 5,162,217 issued Nov. 10, 1992; EP 131,843 published Jan. 23, 1985). Other plasmids involving the N gene, operon, or portion thereof have been described (SU 1405313 published March 15, 1994; EP 314,184 published May 3, 1989; U.S. Pat. No. 4,578,355 issued

March 25, 1986; WO 85/04418 published October 10, 1985; Rees *et al.*, Proc. Natl. Acad. Sci. USA, 93: 342-346 (1996); Hwang *et al.*, Biochem. Biophys. Res. Commun., 173: 711-717 (1990); Bielowski *et al.*, Acta Biochim. Pol., 34: 29-34 (1987); Stanssens *et al.*, Cell, 44: 711-718 (1986); Gatenby and Castleton, Mol. Gen. Genet., 185: 424-429 (1982)); Martin-Gallardo *et al.*, J. Gen. Virol., 74: 453-458 (1993); Das, 72<sup>nd</sup> Annual Meeting of the American Society of Biological Chemists, May 31-June 4, 1981, Fed. Proc., 40 (6): 1764 (1981); Beck *et al.*, Bio/Technology, 6: 930-935 (1988)).

The expression of gamma-interferon was found to increase over two-fold when the  $\lambda$ N anti-termination system was eliminated and only the P<sub>L</sub> promoter was used (WO 85/02624 published June 20, 1985). Cloning and expression vectors in which the active N gene is preferably absent are also described (U.S. Pat. No. 5,401,658 issued March 28, 1995).

Transcription of DNA is often arrested at sites in DNA that trap a fraction of elongating RNA polymerase molecules that pass through, resulting in locked ternary complexes that cannot propagate or dissociate their RNA product. Transcript cleavage factors cleave the RNA in such complexes at the 3' end, allowing RNA polymerase to back up and re-attempt to read through the potential trap. In addition to assuring efficient transcript elongation, transcript cleavage factors increase the fidelity of transcription, since misincorporated bases at the 3' end of the nascent RNA also lead to arrested complexes (Erie *et al.*, Science, 262: 867-873 (1993)). Further, such factors allow RNA polymerase to transcribe through strong blocks to elongation that can otherwise arrest the enzyme on the DNA (Lee *et al.*, J. Biol. Chem., 269: 22295-22303 (1994)). In addition, these factors can facilitate the transition of RNA from the stage of abortive initiation to elongation at certain promoters (Hsu *et al.*, Proc. Natl. Acad. Sci. USA, 92: 11588-11592 (1995)).

Both bacteria and eukaryotes contain proteins that can stimulate such cleavage (Surratt *et al.*, Proc. Natl. Acad. Sci. USA, 88: 7983-7987 (1991); Borukhov *et al.*, Proc. Natl. Acad. Sci. USA, 89: 8899-8902 (1992); Borukhov *et al.*, Cell, 72: 459-466 (1993); Izban and Luse, Genes & Dev., 6: 1342-1356 (1992); Izban and Luse, J. Biol. Chem., 267: 13647-13655 (1992); Izban and Luse, J. Biol. Chem., 268: 12864-12873 (1993); Izban and Luse, J. Biol. Chem., 268: 12874-12885 (1993); Kassavetis and Geiduschek, Science, 259: 944-945 (1993); Reines, J. Biol. Chem., 267: 3795-3800 (1992); Wang and Hawley, Proc. Natl. Acad. Sci. USA, 90: 843-847 (1993); Gu *et al.*, J. Biol. Chem., 268: 25604-25616 (1993); Guo and Price, J. Biol. Chem., 268: 18762-18770 (1993)). Two modes of cleavage have been described. One yields one

to three nucleotide fragments and the other produces larger fragments, up to at least 12 nucleotides in size. Two transcript cleavage factors, GreA and GreB, have been identified in *E. coli* (Borukhov et al., Proc. Natl. Acad. Sci. USA, *supra*, and Borukhov et al., Cell, *supra*, respectively).

5 GreA-dependent transcript cleavage usually results in the removal of di- and trinucleotides from the 3' end of the stalled RNA. GreB-dependent cleavage yields larger oligonucleotides, up to a length of nine nucleotides. Both proteins bind RNA polymerase. Neither the GreA nor GreB proteins possess intrinsic nuclease activity; rather, they stimulate a  
10 nuclease activity inherent in RNA polymerase (Oriova et al., Proc. Natl. Acad. Sci. USA, 92: 4596-4600 (1995)). The GreA and GreB proteins are homologous, sharing 38% sequence identity and 59% sequence similarity. It was found that GreA-induced transcript cleavage in transcription complexes containing *E. coli* RNA polymerase is controlled by multiple factors,  
15 including nascent transcript location and structure (Feng et al., J. Biol. Chem., 269: 22282-22294 (1994)).

Crystallization of GreA has been disclosed (Darst et al., J. Mol. Biol., 242: 582-585 (1994)) as well as its crystal structure (Stebbins et al., Nature, 373: 636-640 (1995)). The organization and functions of domains of GreA and/or GreB have been investigated (Koulich et al., J. Biol. Chem., 272: 7201-7210 (1997); Koulich et al., J. Mol. Biol., 276: 379-389 (1998); Polyakov et al., J. Mol. Biol., 281: 465-473 (1998)).  
20 Moreover, purification and assay procedures for GreA and GreB are reported (Borukhov and Goldfarb, Meth. Enzymol., 274: 315-326 (1996)). Interactions  
25 between RNA polymerase and transcript affect GreA- and GreB-mediated reverse translocation (Feng et al., J. Cellular Biochem. Suppl., 0: 18C, p. 58 (1994)). Both GreA and GreB have been shown to enhance promoter escape (Hsu et al., Proc. Natl. Acad. Sci. USA, 92: 11588-11592 (1995)).

In eukaryotes, the transcription elongation factor TFIIS, otherwise  
30 known as SII (Reines et al., J. Biol. Chem., 264: 10799-10809 (1989); Sluder et al., J. Biol. Chem., 264: 8963-8969 (1989)), is similar to the GreA and GreB proteins in that it stimulates RNA cleavage from the 3' end of RNA in a stalled complex but does not share significant sequence  
35 homology with the GreA and GreB protein (Borukhov et al., Cell, *supra*). TFIIS stimulates either small or large fragment cleavage, depending on reaction conditions and the particular complex examined (Izban and Luse, J. Biol. Chem., 268: 12874-12885 (1993), *supra*; Wang and Hawley, *supra*). Evidence for functional similarity between prokaryotic and eukaryotic

transcription elongation and read-through mechanisms has been found (Mote and Reines, J. Biol. Chem., 273: 16843-16852 (1998)).

Homologs of *E. coli* GreA have been identified. The predicted amino acid sequence encoded by the *Rickettsia prowazekii* greA gene has 50.3% amino acid identity and 66.9% amino acid similarity to *E. coli* GreA (Marks and Wood, Nucl. Acids Res., 20: 3785 (1992)). The deduced amino acid sequence of GreA from *Pseudomonas aeruginosa* exhibits 65.2% identity to its counterpart in *E. coli* K-12 (Lu et al., J. Bacteriol., 179: 3043-3046 (1997)). *Streptococcus pneumoniae* polypeptide GreA has also been disclosed, along with methods of producing the GreA polypeptide by recombinant means and for utilizing GreA or its antagonists for the treatment or diagnosis of infection (EP 838,525 published April 2, 1998). Further, GreA from *Staphylococcus aureus* has also been disclosed, as well as recombinant methods of making it and methods for utilizing it to screen for antibacterial compounds (EP 893,502 published January 27, 1999).

There is a current and continuing need in the art for improving the quality of recombinant protein produced by host cells such that production of truncated forms of the protein, such as 10Sa-tagged material, is minimized or eliminated. There is also a need for higher amounts of full-length protein produced by prokaryotes.

#### Summary of the Invention

Accordingly, the present invention provides, in one aspect, a vector for producing a polypeptide heterologous to prokaryotic cells comprising (1) anti-termination nucleic acid that inhibits intragenic transcription termination with a non-lambda promoter therefor, and (2) RNA encoding the polypeptide with a non-lambda promoter therefor, wherein an RNA recognition site for binding anti-termination protein produced from the nucleic acid is located 5' of the RNA encoding the polypeptide. Preferably, the vector further comprises nucleic acid encoding a GreA or GreB protein with a promoter therefor.

In another aspect, the invention provides a process for producing a heterologous polypeptide in prokaryotic host cells comprising:

(a) culturing the host cells, which comprise (1) anti-termination nucleic acid that inhibits intragenic transcription termination with a non-lambda promoter therefor, and (2) RNA encoding the polypeptide with a non-lambda promoter therefor, wherein an RNA recognition site for binding anti-termination protein produced from the nucleic acid is located 5' of the RNA encoding the polypeptide, and wherein the anti-termination nucleic acid is expressed at the time of expression of the RNA; and

(b) recovering the heterologous polypeptide from the cells or from cell culture medium.

The invention supplies, in yet another aspect, a vector comprising nucleic acid encoding GreA or GreB protein and nucleic acid encoding a polypeptide heterologous to prokaryotic cells, preferably with one or more promoters for the nucleic acids.

In a still further aspect, the invention entails a process for producing a heterologous polypeptide in prokaryotic host cells comprising:

(a) culturing the host cells, which comprise nucleic acid encoding GreA or GreB protein and nucleic acid encoding the heterologous polypeptide, and one or more promoters for the nucleic acids; and

(b) recovering the heterologous polypeptide from the cells or from cell culture medium.

Lambda phage uses N anti-termination to control gene expression by transcribing through strategically placed intergenic terminators. The anti-termination system herein is found to be effective against intragenic termination signals within heterologous genes.

Moreover, the general trend in the literature over the years is to eliminate the lambda N anti-termination system that was used with the adjacent  $P_L$  promoter system and just use the  $P_L$  promoter or replace the N gene with a polylinker or other fusion partner. In contrast, the invention herein lies in using the anti-termination system without the  $P_L$  promoter.

Further, the system herein is designed specifically to prevent the formation and accumulation of truncated and 10Sa-tagged heterologous proteins, which cause problems with protein purification. One of the primary causes of truncated and 10Sa-tagged proteins is premature transcription termination within a translated coding sequence. The accumulation of full-length protein may be similar with or without the anti-termination system, but the accumulation of the truncated forms is significantly reduced by promoting transcriptional readthrough of intragenic termination signals within the protein's coding sequence.

Additionally, unwanted cleavage of the polypeptide is minimized by inclusion of nucleic acids encoding GreA or GreB.

#### Brief Description of the Drawings

Figure 1 shows the construction of plasmid pMP331.

Figure 2 shows the construction of plasmid pMP843.

Figure 3 shows the construction of plasmid pMP871.

Figure 4 shows the construction of plasmid pMP931.

Figure 5 shows the construction of plasmid pMP945.

Figure 6 shows the construction of plasmid pMP951.

Figure 7 shows the construction of plasmid pMP982.

Figure 8 shows the construction of plasmid pMP1016.

5 Figure 9 shows the construction of plasmid pMP1086.

Figure 10 shows the construction of plasmid pMP1099.

Figure 11 shows the construction of plasmid pMP1201.

Figure 12 shows the construction of plasmid pMP1217.

Figure 13 shows the construction of plasmid pJJ142.

10 Figure 14 shows the construction of plasmid pDR1.

Figure 15 shows the construction of plasmid pDR3.

Figures 16A-16C show an analysis of thrombopoietin (TPO) expression under the transcriptional control of the tryptophan (*trp*) and alkaline phosphatase (AP or *phoA*) promoters. On the left are molecular weight markers in kDs. Lanes 1 are the negative control plasmid pBR322 after induction of the *trp* promoter in M9 media, lanes 2 are the *trp* TPO expression plasmid pMP331 after induction of the *trp* promoter, lanes 3 are the negative control plasmid pBR322 after induction of the AP promoter in C.R.A.P. media, and lanes 4 are the AP TPO expression plasmid pMP1099 after induction of the AP promoter. Figure 16A is a Coomassie blue-stained SDS gel of whole cell extracts; Figure 16B is a histidine/horse-radish-peroxidase (HRP)-probed blot of whole cell extracts (the whole cell lysate is separated by SDS-PAGE, transferred to nitrocellulose, and probed with an agent that binds to the polyhis motif on the TPO leader); and Figure 16C is an anti-10Sa polyclonal antibody Western blot of whole cell extracts from TPO induction cultures. The arrows point to full-length TPO.

Figures 17A and 17B respectively show the insertion of a *nutL* site into the *trp* (SEQ ID NO:1) and AP (SEQ ID NO:2) TPO expression constructs. In the Fig. 17A sequence, the *nutL* sequence was inserted at the beginning of the mRNA sequence based on the promoter -10 box. The Fig. 17B sequence shows the insertion of the *nutL* site into the AP promoter construct. Here, the *nut* site is situated further downstream from the mRNA start site.

Figures 18A and 18B show the expression of  $\lambda$ N protein under the transcriptional control of the *tacII* promoter. The start sequence of the  $\lambda$ N protein is shown in each of Figs. 18A and 18B (SEQ ID NO:3). The Fig. 18A nucleotide sequence (SEQ ID NO:4) shows the fusion of the *tacII* promoter to the  $\lambda$ N coding sequence. In this case the *tacII* Shine-Dalgarno sequence has been deleted to provide for reduced  $\lambda$ N translation, and an alternative sequence with lower 16S ribosomal RNA binding is used. The Fig.

18B nucleotide sequence (SEQ ID NO:5) shows the fusion of the *tacII* promoter to the  $\lambda$ N gene using the complete Shine-Dalgarno sequence for high-level N expression.

Figures 19A-19C show an analysis of TPO expression with the *trp* promoter +/-  $\lambda$ N anti-termination. On the left are molecular weight markers in kDs. Lanes 1 are the negative control plasmid pBR322, lanes 2 are the *trp* TPO expression plasmid pMP331, lanes 3 are the *trp* TPO expression plasmid pMP951 with  $\lambda$ N anti-termination and low-level N expression, and lanes 4 are the *trp* TPO expression plasmid pMP1217 with  $\lambda$ N anti-termination and the high-level expression of N. Figure 19A is a Coomassie blue-stained SDS gel of whole cell extracts; Figure 19B is a histidine/horse-radish-peroxidase (HRP)-probed blot of whole cell extracts; and Figure 19C is an anti-10Sa polyclonal Western blot of whole cell extracts. The arrows point to full-length TPO.

Figures 20A-20C show an analysis of TPO expression with the AP promoter +/-  $\lambda$ N anti-termination. On the left are molecular weight markers in kDs. Lanes 1 are the negative control plasmid pBR322, lanes 2 are the AP TPO expression plasmid pMP1099, lanes 3 are the AP TPO expression plasmid pMP1086 with  $\lambda$ N anti-termination and the low-level N expression, and lanes 4 are the AP TPO expression plasmid pMP1201 with  $\lambda$ N anti-termination and the high-level expression of N. Figure 20A is a Coomassie blue-stained SDS gel of whole cell extracts; Figure 20B is a histidine/horse-radish-peroxidase (HRP)-probed blot of whole cell extracts; and Figure 20C is an anti-10Sa polyclonal antibody Western blot of whole cell extracts. The arrows point to full-length TPO.

Figure 21 shows the construction of plasmid pFGF5IT.

Figure 22 shows the construction of plasmid pFGF5IT-AT.

Figures 23A-23C show an analysis of FGF-5 expression +/-  $\lambda$ N anti-termination. Lanes 1 are the negative control plasmid pBR322, lanes 2 are the expression of FGF-5 without  $\lambda$ N anti-termination (pFGF5IT), and lanes 3 are the expression of FGF-5 with  $\lambda$ N anti-termination (pFGF5IT-AT). Fig. 23A is a Coomassie blue-stained SDS gel of whole cell lysates from induced fermentation cultures (equivalent O.D.<sub>600</sub>); Fig. 23B is a Western blot of whole cell lysates probed with an anti-FGF-5 antibody; and Fig. 23C is a Western blot of whole cell lysates probed with an anti-10Sa antibody. The arrows point to full-length FGF-5.



Description of the Preferred Embodiments

As used herein, "*greA*" and "*greB*" refer to genes encoding the transcript cleavage factors known in the literature as GreA and GreB proteins, respectively, and amino acid sequence variants thereof that are functional and effective for the same purpose of transcript cleavage useful in the invention disclosed herein. They may be from any source, with one example being the *greA* and *greB* genes from *E. coli* as described by Borukhov et al., Proc. Natl. Acad. Sci. USA, *supra*, and Borukhov et al., Cell, *supra*, and another example being the *greA* genes disclosed by EP 838,525.

As used herein, "anti-termination factor" is an anti-terminator protein that generally has RNA binding activity and anti-terminator activity. Examples include the anti-terminator N proteins of phages  $\lambda$ ,  $\phi 21$ , and P22, which have been completely sequenced. See Franklin, J. Mol. Biol., 181: 85-91 (1985) and Lazinski et al., Cell, 59: 207-218 (1989). All of these N proteins contain an arginine-rich domain corresponding to about amino acids 1-19 at the N-terminus of the protein that is responsible for RNA binding activity of these proteins, while the remainder of each protein confers anti-terminator activity (Franklin, J. Mol. Biol., 231: 343-360 (1993)). Preferably, the anti-termination factor is a phage factor. More preferably it is a  $\lambda N$  or  $\lambda Q$  gene, more preferably a phage  $\lambda N$  protein.

An "RNA recognition site" refers to a site on an RNA molecule that recognizes a specific protein. For example, an RNA recognition site for binding anti-termination protein would be nut for the N gene, and qut for the Q gene.

A "transcriptional terminator" or "transcriptional termination signal" is operationally defined as a point where the rate of release of an RNA transcript is greater than the rate of addition of the next nucleotide. For purposes herein, the terminator may be rho-dependent or rho-independent. An "intragenic" terminator is one that is homologous to the heterologous polypeptide coding sequence herein. An "intergenic" terminator is one that is exogenous to the heterologous polypeptide coding sequence herein, for example, bacterial or bacteriophage termination signals when the polypeptide is mammalian in origin.

"Anti-termination nucleic acid that inhibits intragenic transcription termination" signifies nucleic acid encoding anti-termination factors that block or override intragenic transcriptional terminators within heterologous genes. This definition includes the N and Q genes, as well as

*rrn* and HK022 anti-termination nucleic acid. Preferably, it is the N or Q gene.

As used herein, "cistron" is a distinctly translatable sequence defined by having a single messenger RNA transcript with one promoter.

5 As used herein, "polycistronic" refers to a polynucleotide comprising two or more cistrons where several different genes are transcribed as a single message from their operons, and two or more pairs of start and stop codons.

10 As used herein, "polypeptide" or "polypeptide of interest" refers generally to peptides and proteins having more than about ten amino acids. The polypeptides are "heterologous," i.e., foreign to the host cell being utilized, such as a human protein produced by a bacterial cell, or a bacterial polypeptide produced from a bacterial cell line that is not the native source of the polypeptide. Preferably, the polypeptide is  
15 mammalian, and most preferably human.

Examples of mammalian polypeptides include molecules such as, e.g., renin, a growth hormone, including human growth hormone; bovine growth hormone; growth hormone releasing factor; parathyroid hormone; thyroid stimulating hormone; lipoproteins;  $\alpha$ 1-antitrypsin; insulin A-chain; insulin  
20 B-chain; proinsulin; thrombopoietin; follicle stimulating hormone; calcitonin; luteinizing hormone; glucagon; clotting factors such as factor VIIIIC, factor IX, tissue factor, and von Willebrands factor; anti-clotting factors such as Protein C; atrial natriuretic factor; lung surfactant; a plasminogen activator, such as urokinase or human urine or tissue-type  
25 plasminogen activator (t-PA); bombesin; thrombin; hemopoietic growth factor; tumor necrosis factor- $\alpha$  and - $\beta$ ; enkephalinase; a serum albumin such as human serum albumin; mullerian-inhibiting substance; relaxin A-chain; relaxin B-chain; prorelaxin; mouse gonadotropin-associated peptide; a microbial protein, such as beta-lactamase; DNASE; inhibin;  
30 activin; vascular endothelial growth factor (VEGF); receptors for hormones or growth factors; integrin; protein A or D; rheumatoid factors; a neurotrophic factor such as brain-derived neurotrophic factor (BDNF), neurotrophin-3, -4, -5, or -6 (NT-3, NT-4, NT-5, or NT-6), or a nerve growth factor such as NGF- $\alpha$ ; cardiostrophins (cardiac hypertrophy factor)  
35 such as cardiostrophin-1 (CT-1); platelet-derived growth factor (PDGF); fibroblast growth factors such as aFGF, bFGF, and FGF-5; epidermal growth factor (EGF); transforming growth factor (TGF) such as TGF- $\alpha$  and TGF- $\beta$ , including TGF- $\beta$ 1, TGF- $\beta$ 2, TGF- $\beta$ 3, TGF- $\beta$ 4, or TGF- $\beta$ 5; insulin-like growth factor-I and -II (IGF-I and IGF-II); des(1-3)-IGF-I (brain IGF-I),

insulin-like growth factor binding proteins; CD proteins such as CD-3, CD-4, CD-8, CD-19, CD-20, and CD-40; erythropoietin; osteoinductive factors; immunotoxins; a bone morphogenetic protein (BMP); an interferon such as interferon-alpha, -beta, and -gamma; colony stimulating factors (CSFs),  
5 e.g., M-CSF, GM-CSF, and G-CSF; interleukins (ILs), e.g., IL-1 to IL-10; anti-HER-2 antibody; superoxide dismutase; T-cell receptors; surface membrane proteins; decay accelerating factor; viral antigen such as, for example, a portion of the AIDS envelope; transport proteins; homing receptors; addressins; regulatory proteins; antibodies; and fragments of  
10 any of the above-listed polypeptides.

Preferred mammalian polypeptides include t-PA, gp120, anti-HER-2, anti-CD20, anti-CD11a, anti-CD18, anti-CD40, DNase, IGF-I, IGF-II, FGF-5, thrombopoietin, brain IGF-I, thrombopoietin, growth hormone, relaxin chains, growth hormone releasing factor, insulin chains or pro-insulin,  
15 urokinase, immunotoxins, neurotrophins, and antigens. Particularly preferred mammalian polypeptides include, e.g., t-PA, gp120(IIIb), anti-HER-2, anti-CD20, anti-CD11a, anti-CD18, anti-CD40, DNase, thrombopoietin, IGF-I, IGF-II, FGF-5, growth hormone, NGF, NT-3, NT-4, NT-5, and NT-6, and most preferably, FGF-5 and thrombopoietin.

20 The expression "control sequences" refers to DNA sequences necessary for the expression of an operably linked coding sequence in a particular host organism. The control sequences that are suitable for prokaryotes include a promoter such as the AP or trp promoter, optionally an operator sequence, and a ribosome-binding site.

25 Nucleic acid is "operably linked" when it is placed into a functional relationship with another nucleic acid sequence. For example, DNA for a presequence or secretory leader is operably linked to DNA for a polypeptide if it is expressed as a preprotein that participates in the secretion of the polypeptide; a promoter is operably linked to a coding sequence if it  
30 affects the transcription of the sequence; or a ribosome binding site is operably linked to a coding sequence if it is positioned so as to facilitate translation. Generally, "operably linked" means that the DNA sequences being linked are contiguous and, in the case of a secretory leader, contiguous and in reading phase. Linking is accomplished by  
35 ligation at convenient restriction sites. If such sites do not exist, the synthetic oligonucleotide adaptors or linkers are used in accordance with conventional practice.

As used herein, the expressions "cell," "cell line," and "cell culture" are used interchangeably and all such designations include

progeny. Thus, the words "transformants" and "transformed cells" include the primary subject cell and cultures derived therefrom without regard for the number of transfers. It is also understood that all progeny may not be precisely identical in DNA content, due to deliberate or inadvertent mutations. Mutant progeny that have the same function or biological activity as screened for in the originally transformed cell are included. Where distinct designations are intended, it will be clear from the context.

A "nut" site refers to the N-utilization site to which the N protein binds. The term includes both natural nut and manipulations or variations thereof that still work with the N protein by binding thereto to effect mRNA-specific lambda N anti-termination. Examples include lambda nutL, nutR, Box B by itself, Box A and Box B, mutant nut sites, and nut sites from related lambdoid phages. These are described, for example, in Friedman et al., Genes Dev., 4: 2210-2222 (1990); Mogridge et al., J. Biol. Chem., 273: 4143-4148 (1998); Olson et al., Cell, 31: 61-70 (1982); Patterson et al., J. Mol. Biol., 236: 217-228 (1994); and Schauer et al., J. Mol. Biol., 194: 679-690 (1987).

A "non-lambda promoter" and "non-lambda" termination sites indicate respectively promoters and termination sites not from lambda phage, for example, not the lambda P<sub>L</sub> promoter. Examples of suitable non-lambda promoters herein include, for example, the *trp* and AP promoters.

#### B. Modes for Carrying out the Invention

In one aspect, a vector is provided for producing a polypeptide heterologous to prokaryotic cells, preferably a mammalian polypeptide, and most preferably a human polypeptide. Such vector has at least the following elements: anti-termination nucleic acid that inhibits intragenic transcription termination, RNA encoding the polypeptide, and one or separate non-lambda promoters for the nucleic acid and RNA. In this vector an RNA recognition site for binding the anti-termination protein produced from the nucleic acid is located 5' of the RNA encoding the polypeptide. The vector may further include nucleic acid encoding a GreA or GreB protein along with a promoter for this nucleic acid, which can be lambda or non-lambda.

In another aspect, a process is described for producing a heterologous polypeptide in prokaryotic host cells. In this process the host cells are cultured and the polypeptide is recovered from the cells or cell culture medium. The cells comprise the components of the above-described vector, i.e., anti-termination nucleic acid that inhibits

intragenic transcription termination and RNA encoding the polypeptide, with non-lambda promoter(s) for each, wherein an RNA recognition site for binding the anti-termination protein produced from the nucleic acid is located 5' of the RNA encoding the polypeptide. In this process the anti-termination nucleic acid is expressed at the time of expression of the RNA.

In another embodiment, a vector is set forth that includes nucleic acid encoding GreA or GreB protein and nucleic acid encoding a heterologous polypeptide, along with one or more promoters therefor, which may be lambda or non-lambda promoters.

In a still further embodiment, a process for producing a heterologous polypeptide in prokaryotic host cells is provided involving culturing the host cells, which comprise the components of the above-described GreA/GreB vector, i.e., nucleic acid encoding GreA or GreB protein and nucleic acid encoding the heterologous polypeptide, and one or more promoters for the nucleic acids. After the culturing step, the heterologous polypeptide is recovered from the cells or from the cell culture medium.

The anti-termination system may be performed with or without the presence of nucleic acid encoding GreA or GreB, since by itself the anti-termination system described above results in a significant decrease in heterologous protein truncation and 10Sa tagging and leads to a corresponding increase in full-length protein. The co-expression with GreA or GreB nucleic acid leads to more recombinant protein production, regardless of whether the anti-termination system described herein is used.

The anti-termination nucleic acid as well as the GreA/GreB nucleic acid and the nucleic acid encoding the heterologous protein may be cDNA or genomic DNA from any source. The anti-termination and GreA/GreB nucleic acids are generally the native sequence, but need not be if they provide the same benefits to heterologous polypeptide production as set forth herein.

If the anti-termination factors or GreA or GreB proteins are native products of the host cell, and if the factors controlling expression of these native genes are understood, such factors can be manipulated to achieve over-expression of these genes, e.g., by induction of transcription from the natural promoter using known inducer molecules, by mutation of the nucleic acids controlling or repressing expression of the gene product to produce a mutant strain that inductively over-expresses the gene product, by second site mutations which depress the synthesis or function of factors that normally repress the transcription of the gene product, and the like.

The heterologous nucleic acid (e.g., cDNA or genomic DNA) is suitably inserted into a replicable vector for expression in the prokaryotic cells under the control of a suitable promoter for prokaryotic cells. Many vectors are available for this purpose, and selection of the appropriate vector will depend mainly on the size of the nucleic acid to be inserted into the vector and the particular host cell to be transformed with the vector. Each vector contains various components depending on its function (amplification of DNA or expression of DNA) and the particular host cell with which it is compatible. The vector components for prokaryotic cell transformation generally include, but are not limited to, one or more of the following: a signal sequence, an origin of replication, one or more marker genes, and a promoter.

The promoters herein may be constitutive or inducible, preferably inducible, and are recognized by the host prokaryotic organism and operably linked to the GreA/GreB-, and/or anti-termination-, and polypeptide-encoding nucleic acid components of the vectors herein. For the anti-termination plasmid, the promoter is non-lambda. For the plasmid that does not contain anti-termination nucleic acid, the promoter may be lambda or non-lambda. The vectors herein contain either one promoter for all two or three elements, provided it is appropriate for all the elements, or two or more separate promoters, which may be the same or different provided they are appropriate, operably linked to each of the nucleic acids encoding the anti-termination factor, the polypeptide, and/or GreA/GreB protein. The promoters are selected to be compatible with the cell type in which expression is to be performed.

For the anti-termination factor the promoter is non-lambda, and for the GreA/GreB the promoter may be lambda or non-lambda. Suitable non-lambda promoters for use in the preferred cell type, *E. coli*, include, for example, the  $\beta$ -lactamase and lactose (*lac*) promoter systems (Chang et al., Nature, 275: 615 (1978); Goeddel et al., Nature, 281: 544 (1979)), the arabinose promoter system (Guzman et al., J. Bacteriol., 174: 7716-7728 (1992)), AP, a *trp* promoter system (Goeddel, Nucleic Acids Res., 8: 4057 (1980) and EP 36,776), hybrid promoters such as the *tac* promoter (deBoer et al., Proc. Natl. Acad. Sci. USA, 80: 21-25 (1983)), and the T3 or T7 promoter (See generally, e.g., Itakura et al., Science, 198: 1056-1063 (1977); Goeddel et al., Proc. Natl. Acad. Sci. USA, 76: 106-110 (1979), Emtage et al., Nature, 283: 171-174 (1980); and Martial et al., Science, 205: 602-606 (1979)). The most preferred non-lambda promoters herein are the *trp* and AP promoters.

Suitable lambda or non-lambda promoters for use with prokaryotic hosts include the non-lambda promoters set forth above, plus the lambda promoters, *i.e.*, those from lambda phage, for example, the lambda P<sub>i</sub> promoter and the lambda P<sub>r</sub> promoter. However, other known lambda and non-lambda promoters are suitable. Their nucleotide sequences have been published, thereby enabling a skilled worker operably to ligate them to DNA encoding the polypeptide of interest or to the anti-termination or *GreA* or *GreB* genes (Siebenlist *et al.*, Cell, 20: 269 (1980)) using linkers or adaptors to supply any required restriction sites.

Promoters for use in prokaryotic systems also generally contain a Shine-Dalgarno (SD) sequence operably linked to the DNA encoding the polypeptide of interest. The promoter can be removed from the prokaryotic source DNA by restriction enzyme digestion and inserted into the vector containing the desired DNA.

If the host contains a *pstS* variant gene, the expression vector for producing a heterologous polypeptide suitably contains an AP promoter that is recognized by the host organism and is operably linked to the nucleic acid encoding the polypeptide of interest. This promoter initiates increased levels of transcription from DNA under its control in response to a decreased concentration of inorganic phosphate in the culture medium. The AP promoter can be removed from the prokaryotic source DNA by restriction enzyme digestion and inserted into the vector containing the desired DNA.

In one alternative, the prokaryotic cells comprise two separate vectors respectively containing the anti-termination nucleic acid and/or *GreA/GreB* protein and the RNA encoding the heterologous polypeptide.

In another alternative, the anti-termination or *greA/greB* nucleic acid and the RNA encoding the heterologous polypeptide are contained on the same vector and are under the control of a single promoter or more than one separate inducible promoters. In this case, the polypeptide may be suitably fused in-frame to an anti-terminator protein and/or *GreA* or *GreB* protein as defined above such that the combined coding sequence is operably linked to a single promoter. Hence, the polypeptide gene and the anti-termination nucleic acid and/or *greA/greB* nucleic acid are suitably coupled to form a polycistronic unit. Alternatively, they may be independently expressed under separate, differently inducible promoters on the same vector so that initiation of expression can occur in the proper order.

The anti-termination and/or *greA/greB* nucleic acid and polypeptide nucleic acid can be anywhere in the cell cytoplasm, including the

chromosome. Hence, they may be integrated into the host cell genome or contained on autonomously replicating plasmids. The anti-termination nucleic acid such as the N or Q gene can be expressed with any reasonably controlled promoter, and is preferably only expressed when the polypeptide RNA is turned on.

The vectors herein that contain the anti-termination nucleic acid also contain a RNA recognition site. This would include the Box B sequence in NutR or NutL for phage  $\lambda$ N protein. This may also include a Box A site if the anti-termination factor is a phage protein. The RNA recognition site is engineered on the polypeptide mRNA at the 5' end. The anti-termination nucleic acid then binds to its RNA recognition site on the polypeptide mRNA and, in conjunction with RNA polymerase and several host factors such as nus factors, in the case of the N gene, forms an anti-termination complex.

Preferably the Nut site is a NutR site, more preferably a  $\lambda$  NutR site, more preferably a complete BoxA and BoxB NutR or a partial BoxB. The complete sequences of Nut sites, which include Box A and Box B domains from phages  $\lambda$ ,  $\phi$ 21, and P22, have been published (Lazinski, Cell, 59: 207-218 (1989)). Box B is responsible for binding an anti-terminator protein. Box A sequences exist not only in phages, but in a variety of other anti-termination operons, including the ribosomal RNA operons of *E. coli* (Friedman and Olson, Cell, 34: 143-149 (1983); Li et al., Cell, 38: 851-860 (1984)). A conserved sequence of 8-12 nucleotides proximal to the promoter in a natural operon, Box A is responsible for binding a host elongation factor that interacts with the anti-termination protein to stimulate anti-termination activity (Greenblatt et al., Nature, 364: 401-406 (1993)).

The Box A domain should preferably match the anti-terminator protein encoded by the gene used. Thus, if the anti-terminator protein is the phage  $\lambda$ N protein, one may choose a  $\lambda$ , NutL, or NutR Box A sequence, which differ slightly in nucleotide sequence. Analogously, if the anti-termination protein is a phage P22 N protein, one may choose a P22 Nut Box A sequence.

The source of the  $\lambda$ N protein is one that is suitable, including from the coding region of the N protein gene of pHE6 (Franklin and Bennett, Gene, 8: 107-119 (1979); EP 467,676 published Jan. 22, 1992) that is removed at the 7 *HinfI* restriction site. Alternatively, one can use the commercially available plasmid from Pharmacia LKB or a plasmid disclosed in EP 700,997, for example.



In general, plasmid vectors containing replicon and control sequences that are derived from species compatible with the host cell are used in connection with prokaryotic hosts. The vector ordinarily carries a replication site, as well as marking sequences that are capable of providing phenotypic selection in transformed cells. For example, *E. coli* is typically transformed using pBR322, a plasmid derived from an *E. coli* species (see, e.g., Bolivar et al., Gene, 2: 95 (1977)). pBR322 contains genes for ampicillin and tetracycline resistance and thus provides easy means for identifying transformed cells. The pBR322 plasmid, or other microbial plasmid or phage, also generally contains, or is modified to contain, promoters that can be used by the microbial organism for expression of the selectable marker genes.

The DNA encoding the polypeptide of interest herein may be expressed not only directly, but also as a fusion with another polypeptide, preferably a signal (leader) sequence or other polypeptide having a specific cleavage site at the N-terminus of the mature polypeptide. In general, the signal sequence may be a component of the vector, or it may be a part of the polypeptide DNA that is inserted into the vector. The heterologous signal sequence selected should be one that is recognized and processed (i.e., cleaved by a signal peptidase) by the host cell. For, e.g., prokaryotic host cells that do not recognize and process the native polypeptide signal sequence, the signal sequence is substituted by a prokaryotic signal sequence selected, for example, from the group consisting of the alkaline phosphatase, penicillinase, lpp, or heat-stable enterotoxin II leader sequences.

The vector also contains a transcription termination site. The choice of termination site is usually not critical. Termination sites are RNA sequences of about 50-100 bases downstream from the translational stop site of a protein-coding sequence. Frequently, RNA termination sites can fold to a hairpin structure. Termination sites are recognized by RNA polymerase as a signal to cease transcription (von Hippel, Science, 255: 809 (1992)). In eukaryotic cells, the selection of termination site depends on the promoter to which the genes are linked. However, in prokaryotic cells, RNA polymerase recognizes virtually any prokaryotic termination site, so the choice of termination site is not critical. In some vectors, multiple termination sites are included in tandem.

Both expression and cloning vectors contain a nucleic acid sequence that enables the vector to replicate in one or more selected host cells. Generally, in cloning vectors this sequence is one that enables the vector

to replicate independently of the host chromosomal DNA, and includes origins of replication or autonomously replicating sequences. Such sequences are well known for a variety of prokaryotes. The origin of replication from the plasmid pBR322 is suitable for most Gram-negative bacteria.

Expression and cloning vectors also generally contain a selection gene, also termed a selectable marker. This gene encodes a protein necessary for the survival or growth of transformed host cells grown in a selective culture medium. Host cells not transformed with the vector containing the selection gene will not survive in the culture medium. Typical selection genes encode proteins that (a) confer resistance to antibiotics or other toxins, e.g., ampicillin, neomycin, methotrexate, or tetracycline, (b) complement auxotrophic deficiencies, or (c) supply critical nutrients not available from complex media, e.g., the gene encoding D-alanine racemase for *Bacilli*. One example of a selection scheme utilizes a drug to arrest growth of a host cell. Those cells that are successfully transformed with a heterologous gene produce a protein conferring drug resistance and thus survive the selection regimen.

Construction of suitable vectors containing one or more of the above-listed components employs standard ligation techniques. Isolated plasmids or DNA fragments are cleaved, tailored, and re-ligated in the form desired to generate the plasmids required.

For analysis to confirm correct sequences in plasmids constructed, the ligation mixtures are used to transform *E. coli* K12 strain 294 (ATCC 31,446) or other strains, and successful transformants are selected by ampicillin or tetracycline resistance where appropriate. Plasmids from the transformants are prepared, analyzed by restriction endonuclease digestion, and/or sequenced by the method of Sanger et al., Proc. Natl. Acad. Sci. USA, 74: 5463-5467 (1977), Messing et al., Nucleic Acids Res., 9: 309 (1981), or Maxam et al., Methods in Enzymology, 65: 499 (1980).

Suitable prokaryotic cells useful as host cells herein include bacteria, for example, archaeobacteria and eubacteria, especially eubacteria, and most preferably *Enterobacteriaceae*. Examples of useful bacteria include *Escherichia*, *Enterobacter*, *Azotobacter*, *Erwinia*, *Bacillus*, *Pseudomonas*, *Klebsiella*, *Proteus*, *Salmonella*, *Serratia*, *Shigella*, *Rhizobia*, *Vitreoscilla*, and *Paracoccus*. These host cells may be lysogenic. Suitable *E. coli* hosts include *E. coli* W3110 (ATCC 27,325), *E. coli* 294 (ATCC 31,446), *E. coli* B, *E. coli* X1776 (ATCC 31,537), and *E. coli* JM105 (New England Biolabs). These examples are illustrative rather than limiting.

Mutant cells of any of the above-mentioned prokaryotic cells may also be employed. It is, of course, necessary to select the appropriate prokaryotic cells taking into consideration replicability of the replicon in the cells of a prokaryote. For example, *E. coli*, *Serratia*, or

5 *Salmonella* species can be suitably used as the host when well-known plasmids such as pBR322, pBR325, pACYC177, or pKN410 are used to supply the replicon.

*E. coli* strain W3110 is a preferred host because it is a common host strain for recombinant DNA product fermentations. Preferably, the host

10 cell should secrete minimal amounts of proteolytic enzymes. For example, strain W3110 may be modified to effect a genetic mutation in the genes encoding proteins, with examples of such hosts including *E. coli* W3110 strain 1A2, which has the complete genotype *tonAΔ* (also known as *ΔfhuA*); *E. coli* W3110 strain 9E4, which has the complete genotype *tonAΔ ptr3*; *E. coli* W3110 strain 27C7 (ATCC 55,244), which has the complete genotype *tonAΔ ptr3 phoAΔE15 Δ(argF-lac)169 ompTΔ degP41kan<sup>r</sup>*; *E. coli* W3110 strain 37D6, which has the complete genotype *tonAΔ ptr3 phoAΔE15 Δ(argF-lac)169 ompTΔ degP41kan<sup>r</sup> rbs7Δ ilvG*; *E. coli* W3110 strain 40B4, which is strain 37D6 with a non-kanamycin resistant *degP* deletion mutation; *E. coli* W3110 strain

20 33D3, which has the complete genotype *tonA ptr3 lacIq LacL8 ompT degP kan<sup>r</sup>*; *E. coli* W3110 strain 36F8, which has the complete genotype *tonA phoA Δ(argF-lac) ptr3 degP kan<sup>r</sup> ilvG+*, and is temperature resistant at 37°C; an *E. coli* strain having the mutant periplasmic protease(s) disclosed in U.S. Pat. No. 4,946,783 issued August 7, 1990; *E. coli* W3110 strain 52A7, which has the complete genotype *tonAΔ (fhuAΔ) lonΔ galE rpoHts (htpRts) ΔclpP lacIq*; *E. coli* W3110 strain 54C2, which has the complete genotype *fhuA(tonA)lon galE rpoHts(htpRts) clpP lacIq*; and *E. coli* W3110 strain 59B9, which has the complete genotype *fhuAΔ(tonAΔ)lonΔ galE rpoHts(htpRts) ΔclpP lacIq ΔompT Δ(nmpc-fepE) ΔlacY*.

30 Transformation means introducing DNA into an organism so that the DNA is replicable, either as an extrachromosomal element or by chromosomal integrant. Depending on the host cell used, transformation is done using standard techniques appropriate to such cells. The calcium treatment employing calcium chloride, as described in section 1.82 of Sambrook et

35 al., Molecular Cloning: A Laboratory Manual (New York: Cold Spring Harbor Laboratory Press, 1989), is generally used for prokaryotic cells that contain substantial cell-wall barriers. Another method for transformation employs polyethylene glycol/DMSO, as described in Chung and Miller, Nucleic

Acids Res., 16: 3580 (1988). Yet another method is the use of the technique termed electroporation.

Host cells are transformed with the above-described expression vectors of this invention and cultured in conventional nutrient media modified as appropriate if promoters are induced. Suitable media for this purpose are described generally, e.g., in Sambrook *et al.*, *supra*. Any other necessary supplements may also be included at appropriate concentrations introduced alone or as a mixture with another supplement or medium such as a complex nitrogen source. The pH of the medium may be any pH from about 5-9, depending mainly on the host organism.

Gene expression may be measured in a sample by any means, including indirectly or directly, for example, by conventional northern blotting to quantitate the transcription of mRNA (Thomas, Proc. Natl. Acad. Sci. USA, 77: 5201-5205 (1980)). Various labels may be employed, most commonly radioisotopes, particularly <sup>32</sup>P. However, other techniques may also be employed, such as using biotin-modified nucleotides for introduction into a polynucleotide. The biotin then serves as the site for binding to avidin or antibodies, which may be labeled with a wide variety of labels, such as radionuclides, fluorescers, enzymes, or the like.

Procedures for observing whether an expressed or over-expressed gene product is secreted are readily available to the skilled practitioner. Once the culture medium is separated from the host cells, for example, by centrifugation or filtration, the gene product can then be detected in the cell-free culture medium or cell culture by taking advantage of known properties characteristic of the gene product. Such properties can include the distinct immunological, enzymatic, or physical properties of the gene product.

For example, if an over-expressed gene product has a unique enzyme activity, an assay for that activity can be performed on the culture medium used by the host cells or extracted cell pellets. Moreover, when antibodies reactive against a given gene product are available, such antibodies can be used to detect the gene product in any known immunological assay (e.g., as in Harlowe *et al.*, Antibodies: A Laboratory Manual (Cold Spring Harbor Laboratory Press: New York, 1988)).

If the gene product is secreted, it can also be detected using tests that distinguish polypeptides on the basis of characteristic physical properties such as molecular weight. To detect the physical properties of the gene product, all polypeptides newly synthesized by the host cell can be labeled, e.g., with a radioisotope. Common radioisotopes that can be

used to label polypeptides synthesized within a host cell include tritium ( $^3\text{H}$ ), carbon-14 ( $^{14}\text{C}$ ), sulfur-35 ( $^{35}\text{S}$ ), and the like. For example, the host cell can be grown in  $^{35}\text{S}$ -methionine or  $^{35}\text{S}$ -cysteine medium, and a significant amount of the  $^{35}\text{S}$  label will be preferentially incorporated into any newly synthesized polypeptide, including the over-expressed heterologous polypeptide. The  $^{35}\text{S}$ -containing culture medium is then removed and the cells are washed and placed in fresh non-radioactive culture medium. After the cells are maintained in the fresh medium for a time and under conditions sufficient to allow secretion of the  $^{35}\text{S}$ -radiolabeled, expressed heterologous polypeptide, the culture medium is collected and separated from the host cells. The molecular weight of the secreted, labeled polypeptide in the culture medium or cell-associated in the cell pellet can then be determined by known procedures, e.g., polyacrylamide gel electrophoresis. Such procedures, and/or other procedures for detecting secreted gene products, are provided, for example, in Goeddel, D.V. (ed.) 1990, *Gene Expression Technology, Methods in Enzymology*, Vol. 185 (Academic Press), and Sambrook et al., *supra*.

For secretion of an expressed or over-expressed gene product, the host cell is cultured under conditions sufficient for secretion of the gene product. Such conditions include, e.g., temperature, nutrient, and cell density conditions that permit secretion by the cell. Moreover, such conditions are those under which the cell can perform basic cellular functions of transcription, translation, and passage of proteins from one cellular compartment to another, as are known to those skilled in the art.

If the secretory elements are in place, the polypeptide of interest is recovered from the periplasm or culture medium as a secreted polypeptide. It is often preferred to purify the polypeptide of interest from recombinant cell proteins or polypeptides and from the anti-termination factor or GreA or GreB protein to obtain preparations that are substantially homogeneous as to the polypeptide of interest. As a first step, the culture medium or lysate may be centrifuged to remove particulate cell debris. The membrane and soluble protein fractions may then be separated if necessary. The polypeptide may then be purified from the soluble protein fraction and from the membrane fraction of the culture lysate, depending on whether the polypeptide is membrane-bound, is soluble, or is present in an aggregated form. The polypeptide thereafter is solubilized and refolded, if necessary, and is purified from contaminant soluble proteins and polypeptides.

One method for isolating exogenous polypeptides from a complex biological mixture containing polypeptides and non-polypeptides contained in a fermentation broth involves contact of reagents with the cells, preferably the cell culture, containing the polypeptide in a non-native conformation, so that an aqueous extraction/isolation can take place. Preferably, the method entails direct addition of reagents to the fermentation vessel after the polypeptide has been produced recombinantly, thereby avoiding extra steps of harvesting, homogenization, and centrifugation to obtain the polypeptide. While the remaining particulates can be removed by Gaulin homogenization and re-suspension, filtration, or a combination thereof, this method utilizes a multiple-phase extraction system for purifying recombinant polypeptides from the remaining particulates.

The following procedures are exemplary of suitable purification procedures: fractionation on immunoaffinity or ion-exchange columns; ethanol precipitation; reverse phase HPLC; chromatography on silica or on a cation-exchange resin such as DEAE; chromatofocusing; SDS-PAGE; ammonium sulfate precipitation; and gel filtration using, for example, SEPHADEX™ G-75 medium from Amersham Biosciences.

The invention will be more fully understood by reference to the following examples. They should not, however, be construed as limiting the scope of the invention. All literature and patent citations mentioned herein are expressly incorporated by reference.

#### EXAMPLE 1

Effect of  $\lambda$ N Anti-termination System on TPO Production (Shake-flask)

##### **Materials and Methods:**

##### **Description and Construction of *E. coli* Expression Vectors**

The plasmids pMP331, pMP843, pMP871, pMP931, pMP951, pMP982, pMP945, pMP1016, pMP1086, pMP1099, pMP1201, and pMP1217 are all designed to express full-length mature TPO (deSavage *et al.*, Nature, 369: 533-538 (1994)) in the *E. coli* cytoplasm using a pBR322-based vector (Bolivar *et al.*, Gene, 2: 95-113 (1977)). The 332-amino-acid TPO coding sequence is preceded in all plasmids by a leader consisting of the first seven amino acids of the heat-stable enterotoxin II signal sequence (Picken *et al.*, Infect. Immun., 42: 269-275(1983)), followed by eight histidine residues, and finally the thrombin cleavage site IEPR (SEQ ID NO:6). Transcription in the plasmids pMP331, pMP843, pMP871, pMP931, pMP945, pMP951, pMP982, and pMP1217 is under the control of the *trp* promoter (Yanofsky *et al.*, Nucleic Acids Res., 9: 6647-6668 (1981)), while pMP1016, pMP1086, pMP1099, and pMP1201 use the

AP promoter (Kikuchi et al., Nucleic Acids Res., 9: 5671-5678 (1981)). Just downstream of the TPO coding sequence is situated the  $\lambda$  transcriptional terminator (Scholtissek et al., Nucleic Acids Res., 15: 3185 (1987)).

5 Additional genetic elements on pMP843, pMP871, pMP931, and pMP945 include the partial  $\lambda$  *nutR* site Box B (or B box), while plasmids pMP951, pMP982, pMP1086, pMP1201, and pMP1217 have the complete  $\lambda$  *nutR* site of Boxes A and B (Olson et al., Cell, 31: 61-70 (1982)) at the site where the TPO-encoding message begins. Plasmids pMP871, pMP931, pMP945, pMP951, pMP982, 10 pMP1086, pMP1201, and pMP1217 also contain the gene for  $\lambda$ N protein (Franklin, J. Mol. Biol., 181: 75-84 (1984)). Finally, plasmids pMP982, pMP1086, pMP1099, and pMP1201 contain the sequence for a rare arginine tRNA, the *argU* gene (*dnaY* gene) (Garcia et al., Cell, 45: 453-459 (1986)).

The plasmid pDR1 is designed to express the protein GreB (Borukhov et 15 al., Cell, *supra*) under the control of the *tacII* promoter (DeBoer et al., *supra*). The backbone of this plasmid is pACYC177 (Chang et al., J. Bacteriol., 134: 1141-1156 (1978); Rose, Nucleic Acids Res., 16: 356 (1988)), which allows it to be compatibly maintained in the same *E. coli* cell with the pBR322-based plasmids.

#### 20 Plasmid pMP331

The plasmid pMP331 is a derivative of the TPO expression plasmid pMP202 (WO 95/18858 published 7/13/95). Briefly, pMP202 is designed to express the first 155 amino acids of TPO downstream of a leader comprising seven amino acids of the STII signal sequence, eight histidines, and a thrombin cleavage site. The plasmid pMP331 extends the TPO coding sequence 25 from 155 to 332 amino acids.

Three DNA fragments were ligated together to make pMP331 as shown in Figure 1, the first of which was the large fragment of the vector pdh108 previously cut with *XbaI* and *StuI*. pdh108 is derived from the vector 30 pGH207-1 (DeBoer et al., Promoters: Structure and Function (Praeger: New York, 1982), pp. 462-481) and contains the  $\lambda$  transcriptional terminator downstream of the *trp* promoter. The second part was the 431-base-pair *XbaI*-*BamHI* fragment from the plasmid pMP202 encoding the first 122 amino acids of TPO. The third part was an approximately 669-base-pair *BamHI*-*RsaI* 35 fragment from phmp11 (deSauvage et al., Nature, 369: 533-538 (1994)) encoding the last half of the TPO gene.

#### Plasmid pMP843

The plasmid pMP843 is the result of adding the  $\lambda$  *nutB* box (*nutR* Box B) downstream of the *trp* promoter in plasmid pMP331. Two fragments were

ligated together to produce pMP843 as shown in Figure 2, the first of which was pMP331 in which the small *SpeI-SacI* fragment had been removed. The second was a 642-base-pair *SpeI-SacI* fragment obtained by ligating a synthetic DNA duplex to the 582-base-pair *XbaI-SacI* fragment from pMP331.

5 The synthetic DNA duplex had the following sequence:

5'-CTAGTTAACTAGTACGCATTCCAGCCCTGAAAAAGGGCAAAGTTCACGTAAAAAGGATAT  
AATTGATCATGCGTAAGGTCGGGACTTTTCCCGTTTCAAGTGCATTTTCTATAGATC-5'  
(SEQ ID NOS:7 and 8, respectively)

10 Plasmid pMP871

The plasmid pMP871 is derived from pMP843 and contains the gene for  $\lambda$ N protein polycistronically coupled downstream of the TPO coding sequence. pMP871 was constructed as shown in Figure 3, by ligating together three DNA fragments, the first of which was the vector pMP843 from which the small *SacI-NheI* fragment had been removed. The second is an approximately 500-base-pair *SacI-FokI* fragment prepared by pre-ligating a synthetic DNA duplex to the 459-base-pair *SacI-AlwNI* fragment from pMP843 encoding amino acids 174-327 of TPO. The synthetic DNA duplex had the following sequence:

15 5'-CTGTCTCAGGAAGGGTAAGCTTTTATGGATGCACAAACAC  
20 TTAGACAGAGTCCTTCCCATTCGAAAATACCTACGTGTTGTGCGGC-5'  
(SEQ ID NOS:9 and 10, respectively)

25 The final part in the ligation was an approximately 657-base-pair *FokI-NheI* fragment from the commercial vector pPL- $\lambda$  (Pharmacia Biotech Inc.) encoding amino acids 7-107 of  $\lambda$ N protein.

Plasmid pMP931

30 The plasmid pMP931 is a derivative of pMP843 in which the  $\lambda$ N gene is placed under the control of the *tacII* promoter. As shown in Figure 4, pMP931 was constructed by ligating together three DNA fragments. The first of these was the vector pMP843 in which the small *ClaI-NheI* fragment had been removed. The second was an approximately 100-base-pair *HinPI-HindIII* fragment from the plasmid pKMTacII (DeBoer et al., (1983), *supra*) containing the *tacII* promoter. The third part in the ligation was a 684-base-pair *HindIII-NheI* fragment from pMP871 encoding the  $\lambda$ N gene.

Plasmid pMP945

40 The plasmid pMP945 is a derivative of pMP931 in which  $\lambda$ N gene expression from the *tacII* promoter is accompanied by a full Shine-Dalgarno sequence for higher translation levels. As shown in Figure 5, this plasmid was constructed by ligating together three DNA fragments, the first of which was the large vector fragment obtained by digesting pMP931 with



*Hind*III and *Nhe*I. The second fragment was a synthetic DNA duplex with the following sequence:

5'-AGCTTAGGATTCTAGAATTATGGATGCACAAACAC  
ATCCTAAGATCTTAATACCTACGTGTTTGTGCGGC- 5'

5 (SEQ ID NOS:11 and 12, respectively)

The last part was the 657-base-pair *Fok*I-*Nhe*I fragment used in the construction of pMP871.

Plasmid pMP951

10 The plasmid pMP951 is derived from pMP931 and additionally contains the full *nut* site (both Boxes A and B). As shown in Figure 6, pMP951 was constructed by ligating together two DNA fragments. The first of these was the vector pMP931 from which the small *Aat*II-*Xba*I fragment had been removed. The second was prepared by pre-ligating a synthetic DNA duplex to the 322-base-pair *Aat*II-*Spe*I fragment from pMP931. The synthetic DNA duplex had the following sequence:

5'-CTAGTTAACTAGTACGCAACGCTCTTACACATTCCAGCC-  
AATTGATCATGCGTTGCGAGAATGTGTAAGGTCGG-

20 -CTGAAAAAGGGCAAAGTTTCACGTAAAAAGGATAT (SEQ ID NO:13)  
-GACTTTTCCCGTTTCAAGTGCATTTTCTCTATAGATC-5' (SEQ ID NO:14)

Plasmid pMP982

25 The plasmid pMP982 is a derivative of pMP951 with the addition of the *argU* gene on the plasmid downstream of the  $\lambda$ N gene. As shown in Figure 7, this plasmid was constructed by ligating together two DNA fragments. The first of these was the vector pMP951 in which the small *Nhe*I-*Sph*I fragment had been removed, and in which the *Nhe*I site had been blunted by treatment with DNA polymerase I (Klenow). The second part was a 440-base-pair *Cla*I-*Sph*I fragment from pST141 in which the *Cla*I site had been blunted by treatment with DNA polymerase I (Klenow). pST141 is a derivative of the plasmid pHGH207-1 (DeBoer et al., (1983), *supra*), and this fragment only encodes the *argU* gene (Garcia et al., Cell, 45: 453-459 (1986)).

Plasmid pMP1016

35 The plasmid pMP1016 is a derivative of pMP331 in which the *trp* promoter has been replaced with the AP promoter. This plasmid was constructed as shown in Figure 8 by ligating together two DNA fragments. The first of these was the vector pST182 in which the small *Xba*I-*Sph*I fragment had been removed. The plasmid pST182 is a derivative of pHGH1 (Chang et al., Gene, 55: 189-196 (1987)), and this latter vector could be used instead to generate this DNA fragment. The second part in the

40

ligation was a 1791-base-pair *XbaI-SphI* fragment from pMP331 encoding the TPO gene and the  $\lambda t_0$  transcriptional terminator.

Plasmid pMP1086

The plasmid pMP1086 is derived from pMP982 and results in the AP promoter being substituted for the *trp* promoter. Three DNA fragments were ligated together to construct pMP1086 as shown in Figure 9. The first of these was the vector pST182 from which the small *SpeI-SphI* fragment had been removed. The plasmid pST182 is a derivative of phGH1 (Chang et al., *Gene, supra*) and this latter vector could be used instead to generate this fragment. The second part in the ligation was a 647-base-pair *SpeI-SacI* fragment from pMP982 encoding the *nut* site and the first 173 amino acids of TPO. The third part was a 1809-base-pair *SacI-SphI* fragment from pMP982 encoding TPO amino acids 174-332, and containing the  $\lambda t_0$  terminator,  $\lambda N$  gene, and *argU* gene.

Plasmid pMP1099

The plasmid pMP1099 is derived from pMP1016 and additionally contains the *argU* gene downstream of the  $\lambda t_0$  transcriptional terminator. As shown in Figure 10, it was constructed by ligating together two DNA fragments. The first of these was the vector pMP1016 in which the small *SacI-SphI* fragment had been removed. The second was a 1020-base-pair *SacI-SphI* fragment from the plasmid pMP591 encoding the last 159 amino acids of TPO, the  $\lambda t_0$  transcriptional terminator, and the *argU* gene. The plasmid pMP591 is a derivative of pMP331 with the addition of the *argU* gene just downstream of the  $\lambda t_0$  transcriptional terminator.

Plasmid pMP1201

The plasmid pMP1201 is a derivative of pMP1086 in which  $\lambda N$  gene expression from the *tacII* promoter is accompanied by a full Shine-Dalgarno sequence for higher translation levels. As shown in Figure 11, this plasmid was constructed by ligating together three DNA fragments, the first of which was the large vector fragment obtained by digesting pMP1086 with *SacI* and *SphI*. The second was a 1007-base-pair *SacI-ClaI* fragment from pMP945 containing the *tacII* promoter with full Shine-Dalgarno and most of the  $\lambda N$  gene. The final piece was a 814-base-pair *ClaI-SphI* fragment obtained by digesting pMP1086 completely with *SphI*, and partially with *ClaI*.

Plasmid pMP1217

The plasmid pMP1217 is a derivative of pMP951 in which  $\lambda N$  gene expression from the *tacII* promoter is accompanied by a full Shine-Dalgarno sequence for higher translation levels. As shown in Figure 12, this plasmid

was constructed by ligating together two DNA fragments. The first of these was the large vector fragment of pMP951 after digestion with *Hind*III and *Nhe*I. The second was a 696-base-pair *Hind*III-*Nhe*I fragment from pMP945 encoding a full Shine-Dalgarno sequence preceding the  $\lambda$ N gene.

5 Plasmid pJJ142

The plasmid pJJ142 is an intermediate in the construction of pDR1 and was prepared by ligating together two DNA fragments as shown in Figure 13. The first of these was the plasmid pACYC177 in which the small *Aat*II-*Hinc*II fragment had been removed. The second part in the ligation was the 1021-  
10 base-pair *Aat*II-*Cla*I fragment from pJJ41 (U.S. Pat. No. 5,639,635) in which the *Cla*I site had been blunted by treatment with DNA polymerase I (Klenow). This latter fragment encodes the *tac*II promoter followed by the *dsbC* gene.

Plasmid pDR1

The plasmid pDR1 is designed to express GreB under the control of the *tac*II promoter. pDR1 was constructed as shown in Figure 14 by the ligation of two DNA fragments. The first of these was the vector pJJ142 in which the small *Xba*I-*Pst*I fragment had been removed. The second part was a 488-  
15 base-pair *Xba*I-*Pst*I fragment containing the *greB* gene. This fragment was prepared by first amplifying the *greB* coding sequence by PCR using *E coli* chromosomal DNA and then digesting this mixture with *Xba*I and *Pst*I. The following primers were used for this step:

5'-CCCCCCCCCTCTAGAAAAATGAAACTCCTCTGGTAACGCGGAAGGG (SEQ ID NO:15)

5'-CCCCCCCCCTGCAGTTACGGTTTCACGTACTCGATAGC (SEQ ID NO:16)

Plasmid pDR3

The plasmid pDR3 is designed to express GreA under the control of the *tac*II promoter. pDR3 was constructed as shown in Figure 15 by the ligation of two DNA fragments. The first of these was the vector pJJ142 in which the small *Xba*I-*Pst*I fragment had been removed. The second part was a 491-  
25 base-pair *Xba*I-*Pst*I fragment containing the *greA* gene. This fragment was prepared by first amplifying the *greA* coding sequence by PCR using *E coli* chromosomal DNA and then digesting this mixture with *Xba*I and *Pst*I. The following primers were used for this step:

5'-CCCCCCCCCTCTAGAATTCTATGCAAGCTATTCGATGACCTTA (SEQ ID NO:17)

5'-CCCCCCCCCTGCAGTTACAGGTATTCCACCTTAAT (SEQ ID NO:18)

35 *E. coli* host

The strain 52A7, which is a derivative of W3110 (ATCC 27,325) having the genotype *tonAΔ (fhuAΔ) lonΔ galE rpoHts (htpRts) ΔclpP lacIq*, was used for the transformation experiments.

Transformation and culturing

For determining if  $\lambda$ N anti-termination had any effects on the truncation and 10Sa tagging of TPO, the *trp* expression vector pMP331 and the anti-termination plasmids with low (pMP951) and high levels (pMP1217) of N expression were transformed into strain 52A7 using standard procedures. All these transformants were grown in Luria Broth (LB) media with ampicillin overnight at 30°C, and then diluted 20 fold into M9 with casamino acids media also containing ampicillin. After approximately 6 hours at 30°C with shaking, the optical density of the cultures (600 nm) was between 2 and 2.5, and isopropyl  $\beta$ -D-thiogalactopyranoside (IPTG) (1mM) was added to the two anti-termination cultures pMP951 and pMP1217. All cultures were then grown an additional 15 hours with shaking at 30°C. Samples were then removed, prepared as described in Yansura *et al.*, Methods in Mol. Biol., 62: 44-62 (1997), and analyzed by SDS-PAGE.

TPO expression with the AP promoter was then tested to see if similar results would be obtained with  $\lambda$ N anti-termination. The AP expression plasmid pMP1099 as well as the anti-termination plasmids with low- (pMP1086) and high-level (pMP1201) N expression were first transformed into the *E. coli* strain 52A7. Transformants were first grown in LB media containing ampicillin overnight at 30°C, and then diluted 100 fold into a phosphate-limiting media called C.R.A.P., which also contains ampicillin. (C.R.A.P. medium contains 3.57g (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 0.71g NaCitrate-2H<sub>2</sub>O, 1.07g KCl, 5.36g yeast extract (Certified), 5.36g HY-CASE® SF refined acid-hydrolyzed casein (Quest International), 110 mL 1M morpholino propane sulfonic acid (MOPS) pH 7.3, 11 mL of a 50% glucose solution, and 7 mL 1M MgSO<sub>4</sub> in a final volume of 1 liter). After growth at 30°C for approximately 6 hours, the cultures reached an optical density (600 nm) of between 1.5 and 2.0, and IPTG (1 mM) was added to the two anti-termination plasmids. Growth at 30°C was continued for another 15 hours, at which time samples were removed, prepared as described for the *trp* vectors, and analyzed by SDS-PAGE.

Results:

Production of TPO using control plasmids pMP331 and pMP1099

TPO (de Sauvage *et al.*, Nature, 369: 533-538 (1994)) was expressed in the cytoplasm of *E. coli* strain 52A7 under the control of both the *trp* (pMP331) and the AP (pMP1099) promoters with a small amino-terminal polyhis leader to provide for easy purification. Induction of either promoter *trp* or AP led to the production of TPO-related protein that could be detected

by Coomassie-stained whole-cell extracts separated on an SDS polyacrylamide gel (SDS-PAGE), as shown in Fig. 16A.

Besides the expected protein band for the full-length polyhis leader-TPO at 37 kD, several other induced protein bands of lower molecular mass were noted. The most noticeable of these had masses of approximately 25, 18, and 14 kD. To determine if these lower-molecular-weight bands were TPO-related, the whole cell lysate was again separated by SDS-PAGE, transferred to nitrocellulose, and probed with an agent that binds to the polyhis motif on the leader (INDIA™ metal-chelated histidine probe bound to HRP (HisProbe-HRP) offered by Pierce Chemical Company) as shown in Fig. 16B. The results verified that the previously noted lower-molecular-weight bands were indeed TPO-related, and also revealed a much more severe truncation problem occurring on the C-terminal end of the protein.

To ascertain if these multiple truncated forms of TPO resulted from degradation of the polyhis leader containing full-length protein or were initially synthesized in this way, a pulse-chase analysis was performed using the same plasmid construct. Without limitation to any one theory, the results suggest that the latter case was the more likely explanation. The formation of truncated forms of TPO can clearly be seen showing up in the very early chase times of 0.5 and 1 minutes and remaining throughout the experiment up to 6 minutes. There was no obvious movement of TPO-related protein from full-length to truncated forms or *vice-versa*, suggesting that the lower-molecular-weight forms were produced directly during protein synthesis.

One possible explanation that needed to be ruled out was plasmid instability, particularly deletions in the TPO-coding sequence. Therefore, the plasmid DNA was isolated by induction culturing, with the resulting DNA subjected to PAGE analysis after digestion with several restriction endonucleases. In addition, individual colonies from the induction culture were analyzed in a similar fashion. In all cases there was no evidence of any plasmid alterations, particularly in the TPO-coding sequence.

Finally, the possibility of producing truncated TPO fragments by the translation of truncated mRNA was investigated. Such protein fragments had previously been shown to contain a C-terminal, 11-amino-acid tag encoded by a small open reading frame in the 10Sa RNA (Tu *et al.*, *supra*; Keiler *et al.*, *supra*). A polyclonal antibody directed against the chemically-synthesized tag was used to probe whole cell extracts from TPO induction cultures. The results in Fig. 16C clearly show that the majority of the truncated TPO fragments contained the 10Sa 11-amino-acid tag, and suggest,

without limitation to any one theory, that the TPO mRNA is truncated or damaged.

Production of TPO using anti-termination system and effect on 10Sa tagging

Although the expression of full-length TPO with the N-terminal polyhis leader was relatively high with both the *trp* and the AP promoters, the accumulation of truncated forms of TPO greatly interfered with protein purification. Since all of the truncated forms have the N-terminal polyhis leader, they co-purify with the full-length protein on a metal chelation column. Elimination of these truncated forms therefore becomes an important factor in the purification and production of a protein such as TPO.

Despite the failure of *rrnB* anti-termination to work for this purpose (Makrides, *supra*), a different transcription anti-termination agent,  $\lambda$ N, was used in this Example to minimize or eliminate the truncation and 10Sa tagging of TPO by promoting transcriptional readthrough of intragenic termination signals within the TPO coding sequence.

Incorporating  $\lambda$ N anti-termination

To incorporate  $\lambda$ N anti-termination into the *trp* and AP expression systems for TPO, the N utilization sequence (*nutL*) was first inserted into the plasmids at locations corresponding to the beginning of the TPO mRNA. The actual design of these sequences, including the upstream promoters and the downstream polyhis leader, is shown in Fig. 17. The  $\lambda$ N gene was then inserted into these plasmids downstream of the TPO coding sequence and the  $\lambda$ t<sub>o</sub> transcriptional terminator. N expression was placed under the control of the *tacII* promoter (DeBoer et al., (1983), *supra*), and plasmids with and without the *tacII* Shine-Dalgarno sequences were constructed. This variation in the translation of N provided an easy way to look at two different expression levels of N, and their subsequent effects on TPO expression. Partial sequences showing the *tacII* promoter with and without the Shine-Dalgarno sequence are shown in Fig. 18.

Effects of  $\lambda$ N anti-termination on TPO expression

Coomassie staining of the SDS gels for the *trp* expression plasmids pMP331, pMP951, and pMP1217 revealed primarily an increase in the expression of full-length TPO with both low-level (pMP951) and high-level (pMP1217) N expression as compared with the control (pMP331) as shown in Fig. 19A. Analysis of the SDS gel, after transferring to nitrocellulose and probing with HisProbe-HRP, showed not only an increase in the expression of full-length TPO, but also a significant decrease in the formation of truncated forms of TPO for both anti-termination plasmids

(Fig. 19B). Finally, an analysis of the gel after transferring to nitrocellulose and probing with antibody to the 10Sa tag also showed a significant decrease in the accumulation of truncated and 10Sa-tagged forms of TPO with both anti-termination plasmids (Fig. 19C).

Coomassie staining of the SDS gels for the AP expression plasmids pMP1099, pMP1086, and pMP1201 showed an increase in the level of full-length TPO with both anti-termination plasmids as compared to the control plasmid pMP1099, as was seen with the *trp* plasmids (Fig. 20A). After transferring to nitrocellulose and probing with HisProbe-HRP, one can also see an increase in the level of full-length TPO as well as a significant decrease in the expression of truncated forms of TPO (Fig. 20B). In a similar analysis, probing with the anti-10Sa antibody showed a decrease in the accumulation of truncated and 10Sa-tagged forms of TPO with both anti-termination plasmids (Fig. 20C).

#### Conclusion:

It is clear from the above shake-flask results that the presence of  $\lambda$ N anti-termination with low or high amounts of N proteins using either promoter increased the titer of the polypeptide. Further, the presence of the  $\lambda$ N protein anti-termination system in low or high amounts with either promoter reduced the amount of 10Sa-tagged protein made by the cells.

### EXAMPLE 2

#### Effect of $\lambda$ N Anti-termination System on TPO Production (Fermentor)

##### Materials and Methods:

##### Transformation

The strain 52A7 was transformed with each of pMP331, pMP951, pMP1099, or pMP1086, each of which is described above in Example 1, using standard procedures involving ampicillin or tetracycline as appropriate.

##### Culture of transformed cells

##### Example for pMP331 and pMP951:

A 10-liter fermentation was carried out in the following medium, with modifications as noted. A bag of salts suitable for a 10-liter fermentation contained the following salts:

<u>Salt</u>	<u>Grams</u>
Ammonium Sulfate	50.0
Potassium Phosphate, dibasic	60.0
Sodium Phosphate, monobasic, dihydrate	30.0
Sodium Citrate, dihydrate	10.0

## Attorney Docket P1732R1

In addition to the salts, 5 g L-isoleucine and 3 mL of a 25% solution of PLURONIC® L-61 antifoam polyol (BASF Corporation) were added to the fermentor. The fermentor was sterilized with these components and 5-6.5 liters of deionized water. After the fermentor and contents cooled down, the post-sterile ingredients were added. The post-sterile ingredients consisted of 15 mL of a 50% glucose solution, 70 mL of 1 M magnesium sulfate, 5 mL of trace metals (recipe below), 250 mL of 20% HY-CASE® acid-hydrolyzed casein solution, 250 mL of 20% yeast extract solution, and 250 mL of a 2 mg/mL ampicillin solution. The starting volume in the fermentor, after inoculation, was usually 8.5 liters.

The fermentor was inoculated with 500 mL of a 16-20-hour LB culture that had been grown with agitation at 30°C. The LB culture was grown in the presence of ampicillin. The 10-liter culture was agitated at 750 rpm and aerated at 10 slpm. The culture pH was maintained at 7.0-7.3 by the automatic addition of ammonium hydroxide, and the temperature was maintained at 30°C. When the initial glucose in the culture was exhausted, a glucose feed was started and maintained at such a rate as to prevent starvation and also to avoid accumulation of glucose in the medium.

Culture growth was monitored by measuring the optical density (O.D.) at a wavelength of 550 nm. When the culture O.D. reached 25-35, 25 mL of a 25 mg/mL solution of 3-β-indole acrylic acid (IAA) and 2-50 mL of a 200 mM IPTG solution (for pMP951 only) were added. The cell paste was harvested via centrifugation 14-18 hours after IAA addition.

<u>Trace Element</u>	<u>Amounts</u>
Hydrochloric Acid	100 mL
Ferric Chloride hexahydrate	27 g/L
Zinc Sulphate heptahydrate	8 g/L
Cobalt Chloride hexahydrate	7 g/L
Sodium Molybdate	7 g/L
Cupric Sulphate pentahydrate	8 g/L
Boric Acid	2 g/L
Manganese Sulphate monohydrate	5 g/L
Deionized Water	to 1 liter

### Example for pMP1099 and pMP1086:

A 10-liter fermentation was carried out in the following medium, with modifications as noted. A bag of salts suitable for a 10-liter fermentation contained the following salts:



# Attorney Docket P1732R1

<u>Salt</u>	<u>Grams</u>
Ammonium Sulfate	50.0
Potassium Phosphate, dibasic	26.0
Sodium Phosphate, monobasic, dihydrate	13.0
5 Sodium Citrate, dihydrate	10.0
Potassium Chloride	15.0

In addition to the salts, 5 g L-isoleucine and 3 mL of a 25% solution of PLURONIC® L-61 antifoam polyol (BASF Corporation) were added to the fermentor. The fermentor was sterilized with these components and 5-6.5 liters of deionized water. After the fermentor and contents cooled down, the post-sterile ingredients were added. The post-sterile ingredients consisted of 15 mL of a 50% glucose solution, 70 mL of 1 M magnesium sulfate, 5 mL of trace metals (recipe below), 250 mL of 20% HY-CASE® acid-hydrolyzed casein solution, 250 mL of 20% yeast extract solution, and 250 mL of a 2 mg/mL-ampicillin solution. The starting volume in the fermentor, after inoculation, was usually 8.5 liters.

The fermentor was inoculated with 500 mL of a 16-20-hour LB culture that had been grown with agitation at 30°C. The LB culture was grown in the presence of ampicillin. The 10-liter culture was agitated at 750 rpm and aerated at 10 slpm. The culture pH was maintained at 7.0-7.3 by the automatic addition of ammonium hydroxide, and the temperature was maintained at 30°C. When the initial glucose in the culture was exhausted, a glucose feed was started and maintained at such a rate as to prevent starvation and also to avoid accumulation of glucose in the medium.

Culture growth was monitored by measuring the O.D. at a wavelength of 550 nm. When the culture O.D. reached 25-35, 2-50 mL of a 200 mM IPTG solution (for pMP1086 only) was added. The cell paste was harvested via centrifugation 20-30 hours after inoculation.

<u>Trace Element</u>	<u>Amounts</u>
Hydrochloric Acid	100 mL
Ferric Chloride hexahydrate	27 g/L
Zinc Sulphate heptahydrate	8 g/L
Cobalt Chloride hexahydrate	7 g/L
35 Sodium Molybdate	7 g/L
Cupric Sulphate pentahydrate	8 g/L
Boric Acid	2 g/L
Manganese Sulphate monohydrate	5 g/L
Deionized Water	to 1 liter

Generation of polyclonal and monoclonal antibodies to the 10Sa peptide:

The following peptide was synthesized for generating antibodies: CAANDENYALAA (SEQ ID NO:19). The N-terminal cysteine is present to allow conjugation of the peptide to either KLH or soybean trypsin inhibitor, or another suitable conjugation partner. The remaining residues are encoded by the *ssrA* gene from *E. coli* and translate to give the 10Sa peptide. The above peptide was synthesized and conjugated to KLH by Zymed Corporation and used as the antigen to raise antibodies in both rabbits and mice.

Specific antibodies to the 10Sa peptide were obtained by taking either serum from rabbits injected with the antigen or ascites fluid from mice. The serum or ascites fluid was passed over an affinity column that had the synthetic 10Sa peptide bound to it. The specific antibodies were eluted from the 10Sa affinity column using low pH.

Quantitation of 10Sa-tagged and full-length TPO:

TPO fusion proteins made from the plasmids pMP331, pMP951, pMP1099, and pMP1086 were extracted from whole cell pellets using a buffer containing 7.5 M guanidine HCl, 0.1 M sodium sulfite, 0.02 M sodium tetrathionate, and 50 mM Tris buffer, pH 8.0. Extractions were allowed to continue 1-16 hours with stirring at room temperature. The extracted solution was then clarified with centrifugation and the supernatant dialyzed 1-16 hours at 4°C against a buffer containing 6 M GuHCl and 20 mM Tris buffer, pH 7.5. The polyhis-containing proteins from the dialysate were purified via a chelating column (Talon Metal Affinity Resins, Clontech). The protein eluted from the column was run on SDS-PAGE, transferred to nitrocellulose, and probed with a polyclonal antibody raised against either TPO<sub>153</sub> (WO 95/18858 published 7/13/95) or the 10Sa peptide. The blots were then scanned using an optically enhanced laser densitometer (PDI Inc., model 325oe). The peak areas for the full-length TPO, as well as the other TPO species that cross-react with the TPO<sub>153</sub> polyclonal antibody, were determined.

Results:

The results from these analyses are shown in Table 1. The ratio of full-length TPO to all of the TPO species was calculated and is reported as % TPO. In addition, the total peak area detected on the blot probed with the polyclonal antibody raised to the 10Sa peptide was also calculated and is reported as 10Sa tag.

Table 1  
Effect of Plasmid Construct on TPO332 Accumulation  
and 10Sa Tagging

	<u>Plasmid</u>	<u>Nut Site</u>	<u>N Protein</u>	<u>% TPO</u>	<u>10Sa Tag</u>
5	pMP331	-	-	4.5	11.5
	pMP951	+	+	49.0	2.6
10	pMP1099	-	-	50.0	4.6
	pMP1086	+	+	73.0	1.6

The data in Table 1 show that expressing TPO from a plasmid with the  $\lambda$ N anti-termination system resulted in an increased percentage of TPO that is full-length TPO and a decrease in accumulated 10Sa tagged TPO. This is seen whether the AP or the *trp* promoter is used to control TPO expression.

### EXAMPLE 3

Effect of GreA or GreB on TPO Production (Shake-flask)

#### Materials and Methods:

##### Transformations

The strain 59B9 (*W3110 fhuAΔ (tonAΔ) lonΔ galE rpoHts(htpRts) ΔclpP lacIq ΔompT Δ(nmpc-fepE) ΔlacY*) was transformed with each of pMP331, pMP951, pMP1217, pMP1099, pMP1086, or pMP1201 either alone or in combination with pDR1 or pDR3, each of which is described above in Example 1, using standard procedures.

##### Culture of transformed cells

The transformed cells were grown in LB media with ampicillin and kanamycin (when co-transformed with pDR1 or pDR3 only) at 30°C with shaking overnight and then diluted 50-fold into shake-flask culture medium containing ampicillin and grown at 30°C with shaking. Transformants containing the plasmids pMP331, pMP951, or pMP1217 were grown in THCD medium with the appropriate antibiotic until they reached an O.D.550 of 1-2, at which time IAA (50 μg/ml final concentration) and IPTG (1 mM, final concentration) (for pMP951 and pMP1217 only) were added to the culture. All cultures were grown for a total of 24 hours. Samples were then removed and prepared for SDS-PAGE.

THCD medium contains 1.86 g Na<sub>2</sub>HPO<sub>4</sub>, 0.93 g NaH<sub>2</sub>PO<sub>4</sub>-H<sub>2</sub>O, 3.57 g (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 0.71 g NaCitrate-2H<sub>2</sub>O, 1.07 g KCl, 5.36 g yeast extract (Difco®

Bacto® brand, #0127-01-7), 5.36 g casamino acids (Difco® Bacto® brand, #0230-17-3), 7 mL of 1M MgSO<sub>4</sub>, 11 mL of a 50% glucose solution, and 110 mL 1 M MOPS, pH 7.3, in a final volume of 1 liter.

Transformants containing the plasmids pMP1099, pMP1086, or pMP1201 with or without either pDR1 or pDR3 were grown in C.R.A.P. medium with the appropriate antibiotics until they reached an O.D.550 of 1-2, at which time IPTG (1 mM final concentration) was added to all of the cultures except for the one containing pMP1099 alone. All cultures were grown for a total of 24 hours. Samples were then removed and prepared for SDS-PAGE.

#### Quantitation of 10Sa-tagged and full-length TPO

Samples from the shake flask cultures described above were prepared and run on SDS-PAGE, transferred to nitrocellulose, and probed with a polyclonal antibody raised against either TPO<sub>153</sub> or the 10Sa peptide. The blots were then scanned using an optically enhanced laser densitometer (PDI, Inc., model 325oe). The peak areas for the full-length TPO, as well as the other TPO species that cross-react with the TPO<sub>153</sub> polyclonal antibody, were determined.

The same set of shake flask samples was also run on SDS-PAGE and stained with Coomassie Blue. The gels were scanned using an optically enhanced laser densitometer (PDI, Inc., model 325oe) and the percentage of total cell protein represented as full-length TPO was calculated.

#### Results:

The results from these analyses are shown in Table 2, where SD signifies the Shine-Dalgarno sequence. The ratio of full-length TPO to all of the TPO species was calculated and is reported as % TPO. In addition, the total peak area detected on the blot probed with the polyclonal antibody raised to the 10Sa peptide was also calculated and is reported as 10Sa tag. The percentage of total cell protein represented as full-length TPO is shown as % of total protein.

Table 2

Effect of Plasmid Construct on TPO332 Accumulation and 10Sa Tagging

	<u>Plasmid(s)</u>	<u>Nut Site</u>	<u>N Protein</u>	<u>% of total protein</u>	<u>%TPO</u>	<u>10Sa Tag</u>
5	pMP331	-	-	5.8	11.3	7.6
10	pMP951	+	+, no SD	5.8	30.9	2.5
	pMP1217	+	+, full SD	5.0	33.0	1.7
15	pMP1099	-	-	6.1	17.8	3.3
	pMP1099/ pDR1	-	-	8.2	21.2	4.2
20	pMP1099/ pDR3	-	-	8.2	21.2	4.8
	pMP1086	+	+, no SD	7.3	27.6	2.3
25	pMP1086/ pDR1	+	+, no SD	7.7	37.5	2.0
	pMP1086/ pDR3	+	+, no SD	8.5	32.0	2.9
30	pMP1201	+	+, full SD	6.8	43.7	1.6

The data in Table 2 show that expressing TPO from a plasmid with the  $\lambda$ N anti-termination system results in a dramatic decrease in 10Sa-tagged TPO, as reflected in the column labeled 10Sa Tag. The percentage of TPO present as full-length TPO also increases with the  $\lambda$ N anti-termination system (% TPO column of Table 2).

It is noted that the control plasmid with the AP promoter (pMP1099), but not with the *trp* promoter (pMP331), produced more % TPO in the fermentor than in the shake flask (compare Table 1 to Table 2). The fermentor conditions provide a more constant glucose feed and controlled pH with a longer induction period than the shake-flask conditions, so that more protein is produced in the former, using the AP promoter. However, with the anti-termination system, the results show a consistent improved trend in %TPO whether the TPO is produced in the shake flasks or in the fermentor and whether the promoter is *trp* or AP. Moreover, the 10Sa tag is lower for the anti-termination system in all cases.

Further, it is clear from the data that co-expressing TPO and either GreA or GreB with or without the  $\lambda$ N anti-termination system results in an

increase in TPO production (% of total protein column in Table 2). There is also an increase in the percentage of TPO present as full-length TPO when either GreA or GreB is co-expressed with TPO.

EXAMPLE 4

Effect of  $\lambda$ N Anti-termination System on FGF-5 Production (Fermentor)

**Materials and Methods:**

Description and Construction of the FGF-5 Expression Plasmids

Plasmid pFGF5IT

The hFGF-5 *E. coli* expression plasmid, pFGF5IT, was constructed from a basic backbone of pBR322 (Sutcliffe, Cold Spring Harb Symp Quant Biol., 43: 77-90 (1978)). The *trp* promoter provides the transcriptional sequence required for efficient expression of the FGF-5 gene in *E. coli* (Yanofsky et al., Nucleic Acids Res., 9: 6647-6668 (1981)). Two Shine-Dalgarno sequences, the *trp* Shine-Dalgarno and a second Shine-Dalgarno, facilitated the translation of FGF-5 mRNA (Yanofsky et al., Nucleic Acids Res., 9: 6647-6668 (1981); Ringquist et al., Molecular Microbiol., 6: 1219-1229 (1992)). The coding sequence for mature FGF-5 (lacking the wild-type signal sequence) is located downstream of the promoter and Shine-Dalgarno sequences and is preceded by a methionine initiation codon.

The vector used for the construction of pFGF5IT was generated by isolating the largest fragment when pRelCIII was digested with *Xba*I and *Bam*HI. This vector contains the *trp* promoter and *trp* Shine-Dalgarno sequence. The second fragment required for this construction was isolated by first digesting pFGF5I with *Hind*III followed by treatment with DNA Polymerase I (Klenow fragment) to create a blunt end. This reaction was then digested with *Xba*I, resulting in a fragment of about 770 bp with one sticky end (*Xba*I) and one blunt end (*Hind*III Pol). This fragment contains a Shine-Dalgarno sequence, an initiation methionine codon, and the coding sequence for mature hFGF-5. The final fragment required for this construction was isolated from pdh108. This *Stu*I-*Bam*HI fragment of about 420 bp contains the sequence encoding the  $\lambda$  transcriptional terminator (Scholtissek et al., Nucleic Acids Res., 15 (7): 3185 (1987)) and approximately the first 375 bp of pBR322 (Sutcliffe, *supra*). These three fragments were ligated together as depicted in Fig. 21 for the construction of pFGF5IT.

Plasmid pFGF5IT-AT

The plasmid pFGF5IT-AT simply places the coding sequence for FGF-5 into a  $\lambda$ N anti-termination expression plasmid. The vector used for this construction was created by isolating the largest fragment when pMP951 was

digested with *Xba*I and *Bam*HI. This vector contains the *trp* promoter and the *nut* site (Boxes A+B). The second fragment required for this construction was isolated following digestion of pFGF5IT-PhoA with *Xba*I and *Hinc*II. The plasmid pFGF5IT-PhoA is a derivative of pFGF5IT in which the *trp* promoter is replaced by the AP promoter (Kikuchi *et al.*, *supra*). This approximately 810-bp fragment contains a Shine-Dalgarno sequence, a methionine initiation codon, the coding sequence for mature hFGF5, and the sequence for the  $\lambda$ t<sub>0</sub> transcriptional terminator. The final fragment required for the ligation was isolated by digestion of pMP931 with *Ssp*I and *Bam*HI. The *Ssp*I digestion was only a partial digestion, resulting in a fragment of approximately 900 bp. This last fragment contains the *tac*II promoter (without a Shine-Dalgarno) followed downstream by the coding sequence for  $\lambda$ N protein. These three fragments were ligated together as illustrated in Fig. 22 to yield the plasmid pFGF5IT-AT.

#### Transformation

The strain 54C2 (*E. coli* W3110 *fhuA(tonA)lon gale rpoHts(htpRts) clpP lacIq*) was transformed with each of pFGF5IT or pFGF5IT-AT using standard procedures involving ampicillin.

#### Culture of transformed cells

The transformed cells were grown up in a fermentor under conditions described in Example 2 for the *trp* plasmids pMP331 and pMP951, except that the IPTG solution was used only for pFGF5IT-AT and not for pFGF5IT. Whole cell lysates from the fermentation samples were prepared for SDS-PAGE.

#### Results:

The FGF-5 intracellular expression plasmid pFGF5IT produced a significant amount of full-length protein (Fig. 23A, Lane 2). Despite this promising result, truncated FGF-5 species also accumulated in addition to the full-length protein (Fig. 23B, Lane 2), and the majority of these species were 10Sa-tagged (Fig. 23C, Lane 2). These results imply, without being limited to any one theory, that premature transcription termination is a likely source of the truncation problem.

To address this problem, the  $\lambda$ N anti-termination system was co-expressed with the FGF-5 gene. The production of full-length FGF-5 was approximately equivalent for both plasmids, with or without  $\lambda$ N anti-termination (Fig. 23A, lanes 2 and 3). However, the accumulation of truncated species was reduced by approximately 50% when the  $\lambda$ N anti-termination system was co-expressed with the FGF-5-encoding gene (Figs. 23B and 23C). The reduction of these truncated species not only allows for simplified purification of the FGF-5 full-length protein, but minimizing

the production of these smaller FGF-5 fragments also leads to improved efficiency in refolding, since the smaller species may contribute to aggregation in the refolding reactions. In addition, any truncated species that do refold and remain in solution may complicate the assessment of bioactivity by interfering with the binding of full-length protein to the receptor. Thus, it is advantageous to reduce the level of premature transcription termination to prevent these potential problems from arising.

#### Example 5

Effects of AN Anti-termination and GreA/B on FGF-5 Production (Shake-flask)

#### Materials and Methods:

##### Shake flask experiments with FGF-5 and GreA or GreB co-expression:

The strain 59B9 {W3110 *fhuAA(tonAA)lonΔgalE rpoHts(htpRts)ΔclpP lacIq ΔompT Δ(nmpc-fepE) ΔlacY*} was transformed with pFGF5IT-PhoA (described in Example 4) or pFGF5IT-PhoAAT, either alone or in combination with pDR1 or pDR3, each of which is described above. The plasmid pFGF5IT-PhoAAT is a derivative of pFGF5IT in which the same anti-termination element is present as in pFGF5IT-AT described in Example 4, and in which the *trp* promoter is replaced by the AP promoter (Kikuchi *et al.*, *supra*).

##### Culture of the transformed cells

The transformed cells were grown in LB media with ampicillin and kanamycin (when co-transformed with pDR1 and pDR3 only) at 30°C with shaking overnight and then diluted 50-fold into C.R.A.P. medium containing the appropriate antibiotics and grown at 30°C with shaking. Transformants were grown in this medium until they reached an OD550 of 1-2, at which time IPTG was added (1 mM, final concentration) to all of the cultures except for the one containing pFGF5IT-PhoA alone. All cultures were grown for a total of 24 hours. Samples were then removed and prepared for SDS-PAGE.

##### Quantitation of 10Sa-tagged and full-length FGF-5:

Samples from the shake flask cultures described above were prepared and run on SDS-PAGE, transferred to nitrocellulose, and probed with a polyclonal antibody raised against either FGF-5 (R&D Systems) or the 10Sa peptide. The blots were then scanned using an optically enhanced laser densitometer (PDI, Inc., model 325oe). The peak areas for the full-length FGF-5, as well as the other FGF-5 species that cross-react with the FGF-5 polyclonal antibody, were determined.

#### Results:

The results from these analyses are shown in Table 3. The ratio of full-length FGF-5 to all of the FGF-5 species was calculated and is reported as % FGF-5. In addition, the total peak area detected on the blot



probed with the polyclonal antibody raised to the 10Sa peptide was also calculated and is reported as 10Sa tag.

Table 3  
Effect of Plasmid Construct on FGF-5 Accumulation and 10Sa Tagging

<u>Plasmid(s)</u>	<u>%FGF-5</u>	<u>10Sa Tag</u>
pFGF5IT-PhoA	15	10.6
pFGF5IT-PhoA/ pDR1	16	11.3
pFGF5IT-PhoA/ pDR3	21	7.6
pFGF5IT-PhoAAT	40	3.0
pFGF5IT-PhoAAT/ pDR1	52	1.7
pFGF5IT-PhoAAT/ pDR3	45	3.4

The data in Table 3 show that expressing FGF-5 from a plasmid with the  $\lambda$ N anti-termination system results in a dramatic decrease in 10Sa-tagged FGF-5, as demonstrated by the scan data for 10Sa-tagged proteins in Table 3. The percentage of FGF-5 present as full-length FGF-5 also increases with the  $\lambda$ N anti-termination system.

Further, it is clear from the data that co-expressing FGF-5 with either GreA or GreB with or without the  $\lambda$ N anti-termination system results in an increase in the percentage of FGF-5 present as full-length FGF-5, as shown by the data in the column labeled %FGF-5.